BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37341
doi: 10.3897/biss.3.37341

OPEN ACCESS

Conference Abstract

Accelerating the Automated Detection, Counting and Measurements of Reproductive Organs in Herbarium Collections in the Era of Deep Learning

Adán Mora-Fallas[‡], Hervé H.G. Goëau[§,|], Susan Mazer[¶], Natalie Love[¶], Erick Mata-Montero[‡], Pierre Bonnet[|,§], Alexis A.J. Joly[#]

‡ School of Computing, Costa Rica Institute of Technology, Cartago, Costa Rica
§ CIRAD, UMR AMAP, Montpellier, France
| AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France
¶ Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, California, United States of America
# Inria, Zenith team, Montpellier, France

Corresponding author: Adán Mora-Fallas (adamora@ic-itcr.ac.cr), Hervé H.G. Goëau (herve.goeau@cirad.fr), Pierre Bonnet (pierre.bonnet@cirad.fr), Alexis A.J. Joly (alexis.joly@inria.fr)

Received: 14 Jun 2019 | Published: 26 Jun 2019

Citation: Mora-Fallas A, Goëau H, Mazer S, Love N, Mata-Montero E, Bonnet P, Joly A (2019) Accelerating the Automated Detection, Counting and Measurements of Reproductive Organs in Herbarium Collections in the Era of Deep Learning. Biodiversity Information Science and Standards 3: e37341. https://doi.org/10.3897/biss.3.37341

Abstract

Millions of herbarium records provide an invaluable legacy and knowledge of the spatial and temporal distributions of plants over centuries across all continents (Soltis et al. 2018). Due to recent efforts to digitize and to make publicly accessible most major natural collections, investigations of ecological and evolutionary patterns at unprecedented geographic scales are now possible (Carranza-Rojas et al. 2017, Lorieul et al. 2019). Nevertheless, biologists are now facing the problem of extracting from a huge number of herbarium sheets basic information such as textual descriptions, the numbers of organs, and measurements of various morphological traits. Deep learning technologies can dramatically accelerate the extraction of such basic information by automating the routines of organ identification, counts and measurements, thereby allowing biologists to spend more time on investigations such as phenological or geographic distribution studies.

© Mora-Fallas A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Recent progress on instance segmentation demonstrated by the Mask-RCNN method is very promising in the context of herbarium sheets, in particular for detecting with high precision different organs of interest on each specimen, including leaves, flowers, and fruits. However, like any deep learning approach, this method requires a significant number of labeled examples with fairly detailed outlines of individual organs. Creating such a training dataset can be very time-consuming and may be discouraging for researchers. We propose in this work to integrate the Mask-RCNN approach within a global system enabling an active learning mechanism (Sener and Savarese 2018) in order to minimize the number of outlines of organs that researchers must manually annotate. The principle is to alternate cycles of manual annotations and training updates of the deep learning model and predictions on the entire collection to process. Then, the challenge of the active learning mechanism is to estimate automatically at each cycle which are the most useful objects that must be manually extracted in the next manual annotation cycle in order to learn, in as few cycles as possible, an accurate model.

We discuss experiments addressing the effectiveness, the limits and the time required of our approach for annotation, in the context of a phenological study of more than 10,000 reproductive organs (buds, flowers, fruits and immature fruits) of *Streptanthus tortuosus*, a species known to be highly variable in appearance and therefore very difficult to be processed by an instance segmentation deep learning model.

Keywords

Herbarium collection, phenology, phenophase, deep learning, instance detection, active learning, visual annotation

Presenting author

Hervé Goëau

Presented at

Biodiversity_Next 2019

References

- Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A (2017) Going deeper in the automated identification of Herbarium specimens. BMC Evolutionary Biology 17: 181. https://doi.org/10.1186/s12862-017-1014-z
- Lorieul T, Pearson KD, Ellwood ER, Goëau H, Molino J-F, Sweeney PW, Yost JM, Sachs J, Mata-Montero E, Nelson G, Soltis PS, Bonnet P, Joly A (2019) Toward a large-scale and

deep phenological stage annotation of herbarium specimens: Case studies from temperate, tropical, and equatorial floras. Applications in Plant Sciences 7 (3): e01233. https://doi.org/10.1002/aps3.1233

- Sener O, Savarese S (2018) Active learning for convolutional neural networks: A core-set approach. International Conference on Learning Representations. URL: https://openreview.net/forum?id=H1aIuk-RW
- Soltis PS, Nelson G, James SA (2018) Green digitization: Online botanical collections data answering real-world questions. Applications in Plant Sciences 6 (2): e1028. https://doi.org/10.1002/aps3.1028